DOMANDE E RISPOSTE MACHINE LEARNING

Nella `parte 1` si trovano le domande senza highlighting delle risposte corrette, mentre nella `parte 2` si trovano le domande comprendenti delle risposte corrette in **grassetto**.

PARTE 1: Domande senza risposte corrette

1) Which is different from the others:

- Expectation Maximisation
- Apriori
- K-means
- Decision Tree

2) What is the cross validation:

- A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set
- A technique to obtain a good estimation of the performance of a classifier with the training set
- A technique to improve the quality of a classifier
- A technique to improve the speed of a classifier

3) What is the single linkage?

- A method to compute the distance between two objects, it can be used in hierarchical clustering
- A method to compute the distance between two sets of items, it can be used in hierarchical clustering
- A method to compute the distance between two classes, it can be used in decision trees
- A method to compute the separation of the objects inside a cluster

4) Which of the following characteristic of data can reduce the effectiveness of DBSCAN?

- Clusters have concavities
- All the variables are the same range of values
- Presence of outliers
- Presence of clusters with different densities

5) Which of the following types of data allows the use of the euclidean distance?

- Points in a vector space
- Ordered data
- Transactional data
- Document representations

6) How does pruning work when generating frequent itemsets?

- If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- If an itemset is frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- If an itemset is not frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated
- If an itemset is frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated

7) In a decision tree, the number of objects in a node...

- ... is smaller than the number of objects in its ancestor
- ... is not related to the number of objects in its ancestor
- ... is smaller than or equal to the number of objects in its ancestor
- ... is bigger than the number of objects in its ancestor

8) What does K-means try to minimise?

- The distortion, that is the sum of the squared distances of each point with respect to the points of the other clusters
- The distortion, that is the sum of the squared distances of each point with respect to its centroid
- The separation, that is the sum of the squared distances of each point with respect to its centroid
- The separation, that is the sum of the squared distances of each cluster centroid with respect to the global centroid of the dataset

9) Given the following definitions: TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives. Which of the formulas below computes the precision of a binary classifier?

- TP / (TP + FP)
- (TP + TN) / (TP + FP + TN + FN)
- TN / (TN + FP)
- TP / (TP + FN)

10) Given the two binary vectors below, which is their similarity according to the Jaccard Coefficient?

1000101101
1011101010

- 0.1
- 0.2
- 0.5
- 0.375

11) Consider the transactional dataset below:

ID Items
1 A,B,C
2 A,B,D
3 B,D,E
4 C,D
5 A,C,D,E

Which is the support of the rule A,C->B

- 100%
- 20%
- 40%
- 50%

12) Which of the following statements is true?

- The noise can generate outliers
- The noise always generate outliers
- The data which are similar to the majority are never noise
- Outliers can be due to noise

13) Which of the following clustering methods is not based on distances between objects?

- Hierarchical Agglomerative
- DBSCAN
- K-Means
- Expectation Maximization

14) In a decision tree, an attribute which is used only in nodes near the leaves...

- ... is irrelevant with respect to the target
- ... guarantees high increment of purity
- ... gives little insight with respect to the target
- ... has a high correlation with respect to the target

15) Which of the statements below is true? (Only one)

- Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters
- K-means always stop to a configuration which gives the minimum distortion for the chosen value of the number of clusters
- K-means works well also with datasets having a very large number of attributes
- K-means finds the number of clusters which give the minimum distortion

16) Which of the statements below is true? (One or more)

- Increasing the radius of the neighourhood can decrease the number of noise points
- DBSCAN always stops to a configuration which gives the optimal number of clusters
- Sometimes DBSCAN stops to a configuration which does not include any cluster
- DBScan can give good performance when clusters have concavities

17) In data preparation which is the effect of normalisation?

- Map all the numeric attributes to the same range, without altering the distribution, in order to avoid that attributes with large ranges have more influence
- Map all the numeric attributes subtracting the respective mean and dividing by variance
- Map all the numeric attributes to the same range and to have a Gaussian (or *normal*) *distribution*, in order to avoid that attributes with large ranges have more influence
- Map all the numeric attributes in order to have a Gaussian (or *normal*) *distribution*, in order to avoid that attributes with large ranges have more influence

18) Consider the transactional dataset below:

ID Items
1 A,B,C
2 A,B,D
3 B,D,E
4 C,D
5 A,C,D,E

Which is the confidence of the rule A,C->B

- 50%
- 40%
- 20%
- 100%

19) Which of the statements below about Hierarchical Agglomerative Clustering is true?

- Requires the definition of Inertia of clusters
- Requires the definition of distance between sets of objects
- Is very efficient also with large datasets
- Is based on a well founded statistical model

20) When training a neural network, what is the learning rate?

- The speed of convergence to a stable solution during the learning process
- The ratio between the size of the hidden layer and the input layer of the network
- The slope of the activation function in a specific node
- A multiplying factor of the correction to be applied to the connection weights

21) What are the hyperparameters of a Neural Network? (Possibly non exhaustive)

- Hidden layers structure, Output layer structure, Activation function, Number of epochs
- Network structure, Learning rate, Backpropagation algorithm, Number of epochs
- Input layers structure, Learning rate, Activation function, Number of epochs
- Hidden layers structure, Learning rate, Activation function, Number of epochs

22) Which of the following preprocessing activities is useful to build a Naive Bayes classifier if the independence hypothesis is violated:

- Discretisation
- Normalisation
- Standardisation
- Feature selection

23) Which is different from the others?

- Silhouette Index
- Misclassification Error
- Entropy
- Gini Index

24) Why do we prune a decision tree?

- To eliminate parts of the tree where the decisions could be influenced by random effects
- To eliminate rows of the dataset which could be influenced by random effects
- To eliminate parts of the tree where the decision could generate underfitting
- To eliminate attributes which could be influenced by random effects

25) In a dataset with D attributes, how many subsets of attributes should be considered for feature selection according ot an exhaustive search?

- $O(2^D)$
- O(D!)
- O(D)
- $O(D^2)$

26) In which mining activity the Information Gain can be useful?

- Discovery of association rules
- Classification
- Clustering
- Discretization

27) Given the two binary vectors below, which is their similarity according to the Simple Matching Coefficient?

1 0 0 0 1 0 1 1 0 1
1 0 1 1 1 0 1 0 1 0

- 0.1
- 0.3
- 0.5
- 0.2

28) Which is the main reason for the standardization of numeric attributes?

- Change the distribution of the numeric attributes, in order to obtain gaussian distributions
- Map all the numeric attributes to a new range such that the mean is zero and the variance is one
- Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence
- Remove non-standard values

29) Which of the following measure can be used as an alternative to the Information Gain?

- Sihlouette Index
- Gini Index
- Rand Index
- Jaccard Index

30) Which of the following is not an objective of feature selection

- Select the features with higher range, which have more influence on the computation
- Reduce time and memory complexity of the mining algorithms
- Reduce the effect of noise
- Avoid the curse of dimensionality

31) After fitting DBSCAN with the default parameter values the results are: 0 clusters, 100% of noise points. Which will be your next trial?

- Reduce the minimum number of objects in the neighborhood
- Reduce the minimum number of objects in the neighborhood and the radius of the neighborhood
- Increase the radius of the neighborhood
- Decrease the radius of the neighborhood

32) What is the meaning of the statement "*the support is anti-monotone*"?

- The support of an itemset never exceeds the support if its subsets
- The support of an itemset is always smaller than the support of its supersets
- The support of an itemsets never exceeds the support if its supersets
- The support of an itemset is always smaller than the support of its subsets

33) In feature selection, what is the Principal Component Analysis?

- A mathematical technique used to transform non numeric attributes into numeric attributes
- A mathematical technique used to transform a set of numeric attributes into a smaller set of numeric attributes which capture most of the variability in data
- A heuristic technique used to find a subset of the attributes which produces the same classifier
- A mathematical technique used to find the principal attributes which determine the classification process

34) A Decision Tree is...

- A tree-structured plan of tests on single attributes to forecast the cluster
- A tree-structured plan of tests on single attributes to forecast the target
- A tree-structured plan of tests on single attributes to obtain the maximum purity of a node
- A tree-structured plan of tests on multiple attributes to forecast the target

35) In data preprocessing, which of the following are the objectives of the aggregation of attributes?

- Obtain a more detailed description of data
- Obtain a less detailed scale
- Reduce the variability of data
- Reduce the number of attributes or distinct values

36) In order to reduce the dimensionality of a dataset, which is the advantage of Multi Dimensional Scaling (MDS), with respect to Principal Component Analysis (PCA)

- MDS requires less computational power
- MDS can be used also with categorical data, provided that the matrix of the distance is available, while PCA is limited to vector spaces
- MDS can be used with categorical data after a transformation in a vector space
- MDS can work on any kind of data, while PCA is limited to categorical data

37) What is the main purpose of smoothing in Bayesian classification?

- Classifying an object containing attribute values which are missing from some classes in the training set
- Dealing with missing values
- Reduce the variability of the data
- Classifying an object containing attribute values which are missing from some classes in the test set

38) Consider the transactional dataset below:

ID Items
1 A,B,C
2 A,B,D
3 B,D,E
4 C,D
5 A,C,D,E

Which is the confidence of the rule B -> E ?

- 100%
- 50%
- 33%
- 20%

39) In a Decision Tree for classification, what is a leaf node?

- A node which assigns a class value to the objects passing the tests on the path from the root to the node itself
- A node where all the objects belong to the same class
- A node which allows classification without errors
- A node which assigns a class value only by majority of examples

40) Match the rule evaluation formulas with their names:

1. $\sup(A \cup C) - \sup(A)\sup(C)$
2. $\frac{\text{conf}(A \Rightarrow C)}{\sup(C)}$
3. $\frac{\sup(A \Rightarrow C)}{\sup(A)}$
4. $\frac{1 - \sup(C)}{1 - \text{conf}(A \Rightarrow C)}$

Options:

- Confidence → ___
- Lift → ___
- Leverage → ___
- Conviction → ___

41) Which of the statements below best describes the strategy of Apriori in finding the frequent itemsets?

- Evaluation of the support of the itemsets in an order such that uninteresting parts of the search space are considered only at the end of the execution
- Evaluation of the confidence of the itemsets in an order such that uninteresting parts of the search space are pruned as soon as possible
- Evaluation of the support of the itemsets in an order such that uninteresting parts of the search space are pruned as soon as possible
- Evaluation of the support of the itemsets in an order such that the interesting parts of the search space are pruned as soon as possible

42) What is the Gini index?

- A measure of the entropy of a dataset
- An accuracy measure of a dataset alternative to the Information Gain and to the Misclassification Index
- An impurity measure of a dataset alternative to the Information Gain and to the Misclassification Index
- An impurity measure of a dataset alternative to overfitting and underfitting

43) What measure is maximised by the Expectation Maximisation algorithm for clustering?

- The likelihood of a class label, given the attributes of the example
- The likelihood of an attribute, given the class label
- The likelihood of an example
- The support of a class

44) Given the following definitions: TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives. Which of the formulas below computes the accuracy of a binary classifier?

- TN / (TN + FP)
- (TP + TN) / (TP + FP + TN + FN)
- TP / (TP + FN)
- TP / (TP + FP)

45) In data preprocessing, which of the following is not an objective of the aggregation of attributes

- Reduce the variability of data
- Obtain a less detailed scale
- Reduce the number of attributes or objects
- Obtain a more detailed description of data

46) Which of the following is a strength of the clustering algorithm DBSCAN?

- Requires to set the number of clusters as a parameter
- Ability to find cluster with concavities
- Ability to separate outliers from regular data
- Very fast by computation

47) Which of the following statements regarding the discovery of association rules is true? (One or more)

- The confidence of a rule can be computed starting from the supports of the itemsets
- The support of an itemset is anti-monotonic with respect to the composition of the itemset
- The support of a rule can be computed given the confidence of the rule
- The confidence of an itemset is anti-monotonic with respect to the composition of the itemset

48) For each type of data choose the best suited distance function

- Vector space with real values -> ___
- Vectors of terms representing documents -> ___
- High dimensional spaces -> ___
- Boolean data -> ___

49) Which of the following is not a strength point of Dbscan with respect to K-means

- The robustness with respect to outliers
- The effectiveness, even in presence of noise
- The effectiveness even if there are clusters with non-convex shape
- The efficiency even in large datasets

50) Which is the effect of the curse of dimensionality

- When the number of dimensions increases the results tend to be prone to overfitting
- When the number of dimensions increases the euclidean distance becomes less effective to discriminate between points in the space
- When the number of dimensions increases the computing power necessary to compute the distances becomes too high
- When the number of dimensions increases the classifiers cannot be correctly tuned

51) Which is different from the others?

- Expectation Maximisation
- Decision Tree
- K-means
- Dbscan

52) In a Neural Network, what is the backpropagation?

- The technique used to adjust the node weights according to the difference between the desired output and the output generated by the network
- The technique used to adjust the weights limiting the probability of overfitting
- The technique used to adjust the output according to the difference between the desired weights and the actual weights
- The technique used to adjust the connection weights according to the difference between the desired output and the output generated by the network

53) Which is different from the others?

- Dbscan
- SVM
- Neural Network
- Decision Tree

54) Which of the following is not a property of a metric distance function

- Symmetry
- Positive definiteness
- Triangle inequality
- Boundedness

55) Which of the statements below is true? (One or more)

- K-mean always stops to a configuration which gives the minimum distortion for the chosen value of the number of clusters
- K-means is very sensitive to the initial assignment of the centers
- K-means is quite efficient even for large datasets
- Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters

56) Which of the following characteristic of data can reduce the effectiveness of K-Means?

- Presence of values with high frequency
- Presence of outliers
- All the variables have the same distribution of values
- All the variables have the same range of values

57) Which is the purpose of discretisation?

- Increase the number of distinct values in an attribute, in order to put in evidence possible patterns and regularities
- Reduce the number of distinct values in an attribute, in order to increase the efficiency of the computations
- Reduce the number of distinct valuess in an attribute, in order to put in evidence possible patterns and regularities
- Reduce the range of values of a numeric attribute, to make all the attributes more comparable

58) Given the following definitions: TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives. Which of the formulas below computes the recall of a binary classifier?

- TP / (TP + FN)
- TN / (TN + FP)
- TP / (TP + FP)
- (TP + TN) / (TP + FP + TN + FN)

59) The Information Gain is used to

- select the attribute which maximises, for a given training set, the ability to predict the class value
- select the attribute which maximises, for a given test set, the ability to predict the class value
- select the attribute which maximises, for a given training set, the ability to predict all the other attribute values
- select the class with maximum probability

60) Which is the main reason for the MinMax scaling (also known as "rescaling") of attributes?

- Change the distribution of the numeric attributes, in order to obtain gaussian distributions
- Map all the nominal attributes to the same range in order to prevent the values with higher frequency from having prevailing influence
- Remove abnormal values
- Map all the numeric attributes to the same range, in order to prevent attributes with higher range from having prevalent influence

61) Which of the following is a base hypothesis for a bayesian classifier?

- The attributes must be statistically independent inside each class
- The attributes must have negative correlation
- The attributes must be statistically independent
- The attributes must have zero correlation

62) When developing a classifier, which of the following is a symptom of overfitting?

- The error rate in the test set is much smaller than the error rate in the training set
- The precision is much greater than the recall
- The error rate in the test set is more than 30%
- The error rate in the test set is much greater than the error rate in the training set

63) With reference to the total sum of squared errors and separation of a clustering scheme, which of the statements below is true?

- They are two ways to measure the same thing
- They are strictly correlated, if, changing the clustering scheme, one increases, then the other does the same
- It is possible to optimise them (i.e. minimise SSE and maximise SSB) separately
- They are strictly correlated, if, changing the clustering scheme, one increases, then the other decreases

64) How can we measure the quality of a trained regression model?

- With a formula elaborating the difference between the forecast values and the true ones
- With a confusion matrix
- Counting the number of values correctly forecast
- With precision, recall and accuracy

65) What is the coefficient of determination R2?

- Provide an index of goodness for a linear regression model
- Measure the amount of error in a linear regression model
- Measure the amount of error in a regression model
- An index of goodness for a classification model

66) What measure is maximised by the Expectation Masimisation algirithm for clustering?

Scegli un'alternativa:

- The likelihood the distributions, defined by the parameters found, given the data available
- The likelihood of an example
- The support of a class
- The likelihood of an attribute, given the class label

67) What is the difference between classification and regression?

- Classification has a categorical target, while regression has a numeric target
- Classification is a supervised activity. while regression in unsupervised
- Classification can make errors, while regression is always exact
- Classification can have a numeric or categorical target, while regression has always a categorical target

68) When is polynomial regression appropriate?

- When the relationship between the predicting variable and the target cannot be approximated as linear
- When the target values are not linearly separable
- When there is more than one predicting attribute
- When it is necessary to project the data into a dimensional space

69) How is the number of clusters determined in agglomerative clustering?

- By assigning each data point to a cluster based on the final structure obtained from the dendrogram.
- By cutting the dendrogram at a certain height.
- By merging the closest clusters according to the defined distance metric.
- By computing the pairwise similarity or distance between all clusters.

70) In data preprocessing, which of the operations below performs aggregation?

- Combination of two attributes in order to obtatn a more general attribute with smaller variability
- Reduction of the number of distinct values of an attribute. in order to reduce its variability
- Reduction of the numbet of rows of the dataset by applying a groupinq on one or more attributes and computing the measures as the results of aggregation functions
- Combination of two or more datasets in Order to obtatn more Information

71) Match specific situations of data to the most appropriate pre-processing activity

- The numeric attributes have very diverse ranges of values : ___
- An attribute has a very skewed distribution (e.g. most of the values concentrated in a small range, and a very long tail of outliers: ___
- A categorical attribute has a too large number of distinct values: ___
- A nominal attribute needs to be used as predicting when using the sklearn library: ___

Alternatives:

- Aggregation
- Ordinal Encoding
- Rescaling
- Non linear transformations
- OneHot Encoding

72) What is the first step in agglomerative clustering?

Scegli un'alternativa:

- Cutting the Dendrogram: Decide on the number of clusters by cutting the dendrogram at a certain height
- Pairwise Cluster Similarity Calculation: Compute the pairwise similarity or distance between all clusters
- Initialization: Start with each data point as a singleton cluster
- Cluster Assignment: Assign each data point to a cluster based on the final structure obtained from the dendrogram

73) Which of the following situations could suggest the use of MinMax scaling? (also known as "rescaling')

- There are numeric attributes with very diverse value ranges
- There are attributes with very skewed data distributions
- There are attrtbutes with a very large number of missing values
- There is a numeric attribute with values spread in a very large range

PARTE 2: Domande con risposte corrette (in grassetto)

1) Which is different from the others:

- Expectation Maximisation
- Apriori
- K-means
- **Decision Tree**

2) What is the cross validation:

- **A technique to obtain a good estimation of the performance of a classifier when it will be used with data different from the training set**
- A technique to obtain a good estimation of the performance of a classifier with the training set
- A technique to improve the quality of a classifier
- A technique to improve the speed of a classifier

3) What is the single linkage?

- A method to compute the distance between two objects, it can be used in hierarchical clustering
- **A method to compute the distance between two sets of items, it can be used in hierarchical clustering**
- A method to compute the distance between two classes, it can be used in decision trees
- A method to compute the separation of the objects inside a cluster

4) Which of the following characteristic of data can reduce the effectiveness of DBSCAN?

- Clusters have concavities
- All the variables are the same range of values
- Presence of outliers
- **Presence of clusters with different densities**

5) Which of the following types of data allows the use of the euclidean distance?

- **Points in a vector space**
- Ordered data
- Transactional data
- Document representations

6) How does pruning work when generating frequent itemsets?

- **If an itemset is not frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated**
- If an itemset is frequent, then none of its supersets can be frequent, therefore the frequencies of the supersets are not evaluated
- If an itemset is not frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated
- If an itemset is frequent, then none of its subsets can be frequent, therefore the frequencies of the subsets are not evaluated

7) In a decision tree, the number of objects in a node...

- **... is smaller than the number of objects in its ancestor**
- ... is not related to the number of objects in its ancestor
- ... is smaller than or equal to the number of objects in its ancestor
- ... is bigger than the number of objects in its ancestor

8) What does K-means try to minimise?

- The distortion, that is the sum of the squared distances of each point with respect to the points of the other clusters
- **The distortion, that is the sum of the squared distances of each point with respect to its centroid**
- The separation, that is the sum of the squared distances of each point with respect to its centroid
- The separation, that is the sum of the squared distances of each cluster centroid with respect to the global centroid of the dataset

9) Given the following definitions: TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives. Which of the formulas below computes the precision of a binary classifier?

- **TP / (TP + FP)**
- (TP + TN) / (TP + FP + TN + FN)
- TN / (TN + FP)
- TP / (TP + FN)

10) Given the two binary vectors below, which is their similarity according to the Jaccard Coefficient?

1000101101
1011101010

- 0.1
- 0.2
- 0.5
- **0.375**

11) Consider the transactional dataset below:

ID Items
1 A,B,C
2 A,B,D
3 B,D,E
4 C,D
5 A,C,D,E

Which is the support of the rule A,C->B

- 100%
- **20%**
- 40%
- 50%

12) Which of the following statements is true?

- **The noise can generate outliers**
- The noise always generate outliers
- The data which are similar to the majority are never noise
- **Outliers can be due to noise**

13) Which of the following clustering methods is not based on distances between objects?

- Hierarchical Agglomerative
- DBSCAN
- K-Means
- **Expectation Maximization**

14) In a decision tree, an attribute which is used only in nodes near the leaves...

- ... is irrelevant with respect to the target
- ... guarantees high increment of purity
- **... gives little insight with respect to the target**
- ... has a high correlation with respect to the target

15) Which of the statements below is true? (Only one)

- **Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters**
- K-means always stop to a configuration which gives the minimum distortion for the chosen value of the number of clusters
- K-means works well also with datasets having a very large number of attributes
- K-means finds the number of clusters which give the minimum distortion

16) Which of the statements below is true? (One or more)

- **Increasing the radius of the neighourhood can decrease the number of noise points**
- DBSCAN always stops to a configuration which gives the optimal number of clusters
- **Sometimes DBSCAN stops to a configuration which does not include any cluster**
- **DBScan can give good performance when clusters have concavities**

17) In data preparation which is the effect of normalisation?

- **Map all the numeric attributes to the same range, without altering the distribution, in order to avoid that attributes with large ranges have more influence**
- Map all the numeric attributes subtracting the respective mean and dividing by variance
- Map all the numeric attributes to the same range and to have a Gaussian (or *normal*) *distribution*, in order to avoid that attributes with large ranges have more influence
- Map all the numeric attributes in order to have a Gaussian (or *normal*) *distribution*, in order to avoid that attributes with large ranges have more influence

18) Consider the transactional dataset below:

ID Items
1 A,B,C
2 A,B,D
3 B,D,E
4 C,D
5 A,C,D,E

Which is the confidence of the rule A,C->B

- **50%**
- 40%
- 20%
- 100%

19) Which of the statements below about Hierarchical Agglomerative Clustering is true?

- Requires the definition of Inertia of clusters
- **Requires the definition of distance between sets of objects**
- Is very efficient also with large datasets
- Is based on a well founded statistical model

20) When training a neural network, what is the learning rate?

- The speed of convergence to a stable solution during the learning process
- The ratio between the size of the hidden layer and the input layer of the network
- The slope of the activation function in a specific node
- **A multiplying factor of the correction to be applied to the connection weights**

21) What are the hyperparameters of a Neural Network? (Possibly non exhaustive)

- Hidden layers structure, Output layer structure, Activation function, Number of epochs
- Network structure, Learning rate, Backpropagation algorithm, Number of epochs
- Input layers structure, Learning rate, Activation function, Number of epochs
- **Hidden layers structure, Learning rate, Activation function, Number of epochs**

22) Which of the following preprocessing activities is useful to build a Naive Bayes classifier if the independence hypothesis is violated:

- Discretisation
- Normalisation
- Standardisation
- **Feature selection**

23) Which is different from the others?

- **Silhouette Index**
- Misclassification Error
- Entropy
- Gini Index

24) Why do we prune a decision tree?

- **To eliminate parts of the tree where the decisions could be influenced by random effects**
- To eliminate rows of the dataset which could be influenced by random effects
- To eliminate parts of the tree where the decision could generate underfitting
- To eliminate attributes which could be influenced by random effects

25) In a dataset with D attributes, how many subsets of attributes should be considered for feature selection according ot an exhaustive search?

- **$O(2^D)$**
- O(D!)
- O(D)
- $O(D^2)$

26) In which mining activity the Information Gain can be useful?

- Discovery of association rules
- **Classification**
- Clustering
- Discretization

27) Given the two binary vectors below, which is their similarity according to the Simple Matching Coefficient?

1 0 0 0 1 0 1 1 0 1
1 0 1 1 1 0 1 0 1 0

- 0.1
- 0.3
- **0.5**
- 0.2

28) Which is the main reason for the standardization of numeric attributes?

- Change the distribution of the numeric attributes, in order to obtain gaussian distributions
- **Map all the numeric attributes to a new range such that the mean is zero and the variance is one**
- Map all the nominal attributes to the same range, in order to prevent the values with higher frequency from having prevailing influence
- Remove non-standard values

29) Which of the following measure can be used as an alternative to the Information Gain?

- Sihlouette Index
- **Gini Index**
- Rand Index
- Jaccard Index

30) Which of the following is not an objective of feature selection

- **Select the features with higher range, which have more influence on the computation**
- Reduce time and memory complexity of the mining algorithms
- Reduce the effect of noise
- Avoid the curse of dimensionality

31) After fitting DBSCAN with the default parameter values the results are: 0 clusters, 100% of noise points. Which will be your next trial?

- **Reduce the minimum number of objects in the neighborhood**
- Reduce the minimum number of objects in the neighborhood and the radius of the neighborhood
- **Increase the radius of the neighborhood**
- Decrease the radius of the neighborhood

32) What is the meaning of the statement "*the support is anti-monotone*"?

- **The support of an itemset never exceeds the support if its subsets**
- The support of an itemset is always smaller than the support of its supersets
- The support of an itemsets never exceeds the support if its supersets
- The support of an itemset is always smaller than the support of its subsets

33) In feature selection, what is the Principal Component Analysis?

- A mathematical technique used to transform non numeric attributes into numeric attributes
- **A mathematical technique used to transform a set of numeric attributes into a smaller set of numeric attributes which capture most of the variability in data**
- A heuristic technique used to find a subset of the attributes which produces the same classifier
- A mathematical technique used to find the principal attributes which determine the classification process

34) A Decision Tree is...

- A tree-structured plan of tests on single attributes to forecast the cluster
- **A tree-structured plan of tests on single attributes to forecast the target**
- A tree-structured plan of tests on single attributes to obtain the maximum purity of a node
- A tree-structured plan of tests on multiple attributes to forecast the target

35) In data preprocessing, which of the following are the objectives of the aggregation of attributes?

- Obtain a more detailed description of data
- **Obtain a less detailed scale**
- **Reduce the variability of data**
- **Reduce the number of attributes or distinct values**

36) In order to reduce the dimensionality of a dataset, which is the advantage of Multi Dimensional Scaling (MDS), with respect to Principal Component Analysis (PCA)

- MDS requires less computational power
- **MDS can be used also with categorical data, provided that the matrix of the distance is available, while PCA is limited to vector spaces**
- MDS can be used with categorical data after a transformation in a vector space
- MDS can work on any kind of data, while PCA is limited to categorical data

37) What is the main purpose of smoothing in Bayesian classification?

- **Classifying an object containing attribute values which are missing from some classes in the training set**
- Dealing with missing values
- Reduce the variability of the data
- Classifying an object containing attribute values which are missing from some classes in the test set

38) Consider the transactional dataset below:

ID Items
1 A,B,C
2 A,B,D
3 B,D,E
4 C,D
5 A,C,D,E

Which is the confidence of the rule B -> E ?

- 100%
- 50%
- **33%**
- 20%

39) In a Decision Tree for classification, what is a leaf node?

- **A node which assigns a class value to the objects passing the tests on the path from the root to the node itself**
- A node where all the objects belong to the same class
- A node which allows classification without errors
- A node which assigns a class value only by majority of examples

40) Match the rule evaluation formulas with their names:

1. $\sup(A \cup C) - \sup(A)\sup(C)$
2. $\frac{\text{conf}(A \Rightarrow C)}{\sup(C)}$
3. $\frac{\sup(A \Rightarrow C)}{\sup(A)}$
4. $\frac{1 - \sup(C)}{1 - \text{conf}(A \Rightarrow C)}$

Options:

- Confidence → 3
- Lift → 2
- Leverage → 1
- Conviction → 4

41) Which of the statements below best describes the strategy of Apriori in finding the frequent itemsets?

- Evaluation of the support of the itemsets in an order such that uninteresting parts of the search space are considered only at the end of the execution
- Evaluation of the confidence of the itemsets in an order such that uninteresting parts of the search space are pruned as soon as possible
- **Evaluation of the support of the itemsets in an order such that uninteresting parts of the search space are pruned as soon as possible**
- Evaluation of the support of the itemsets in an order such that the interesting parts of the search space are pruned as soon as possible

42) What is the Gini index?

- A measure of the entropy of a dataset
- An accuracy measure of a dataset alternative to the Information Gain and to the Misclassification Index
- **An impurity measure of a dataset alternative to the Information Gain and to the Misclassification Index**
- An impurity measure of a dataset alternative to overfitting and underfitting

43) What measure is maximised by the Expectation Maximisation algorithm for clustering?

- **The likelihood of a class label, given the attributes of the example**
- The likelihood of an attribute, given the class label
- The likelihood of an example
- The support of a class

44) Given the following definitions: TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives. Which of the formulas below computes the accuracy of a binary classifier?

- TN / (TN + FP)
- **(TP + TN) / (TP + FP + TN + FN)**
- TP / (TP + FN)
- TP / (TP + FP)

45) In data preprocessing, which of the following is not an objective of the aggregation of attributes

- Reduce the variability of data
- Obtain a less detailed scale
- Reduce the number of attributes or objects
- **Obtain a more detailed description of data**

46) Which of the following is a strength of the clustering algorithm DBSCAN?

- Requires to set the number of clusters as a parameter
- **Ability to find cluster with concavities**
- **Ability to separate outliers from regular data**
- Very fast by computation

47) Which of the following statements regarding the discovery of association rules is true? (One or more)

- **The confidence of a rule can be computed starting from the supports of the itemsets**
- **The support of an itemset is anti-monotonic with respect to the composition of the itemset**
- The support of a rule can be computed given the confidence of the rule
- The confidence of an itemset is anti-monotonic with respect to the composition of the itemset

48) For each type of data choose the best suited distance function

- Vector space with real values -> **euclidean distance**
- Vectors of terms representing documents -> **cosine distance**
- High dimensional spaces -> **manhattan distance**
- Boolean data -> **jaccard coefficient**

49) Which of the following is not a strength point of Dbscan with respect to K-means

- The robustness with respect to outliers
- The effectiveness, even in presence of noise
- The effectiveness even if there are clusters with non-convex shape
- **The efficiency even in large datasets**

50) Which is the effect of the curse of dimensionality

- When the number of dimensions increases the results tend to be prone to overfitting
- **When the number of dimensions increases the euclidean distance becomes less effective to discriminate between points in the space**
- When the number of dimensions increases the computing power necessary to compute the distances becomes too high
- When the number of dimensions increases the classifiers cannot be correctly tuned

51) Which is different from the others?

- Expectation Maximisation
- **Decision Tree**
- K-means
- Dbscan

52) In a Neural Network, what is the backpropagation?

- The technique used to adjust the node weights according to the difference between the desired output and the output generated by the network
- The technique used to adjust the weights limiting the probability of overfitting
- The technique used to adjust the output according to the difference between the desired weights and the actual weights
- **The technique used to adjust the connection weights according to the difference between the desired output and the output generated by the network**

53) Which is different from the others?

- **Dbscan**
- SVM
- Neural Network
- Decision Tree

54) Which of the following is not a property of a metric distance function

- Symmetry
- Positive definiteness
- Triangle inequality
- **Boundedness**

55) Which of the statements below is true? (One or more)

- K-mean always stops to a configuration which gives the minimum distortion for the chosen value of the number of clusters
- **K-means is very sensitive to the initial assignment of the centers**
- **K-means is quite efficient even for large datasets**
- **Sometimes k-means stops to a configuration which does not give the minimum distortion for the chosen value of the number of clusters**

56) Which of the following characteristic of data can reduce the effectiveness of K-Means?

- Presence of values with high frequency
- **Presence of outliers**
- All the variables have the same distribution of values
- All the variables have the same range of values

57) Which is the purpose of discretisation?

- Increase the number of distinct values in an attribute, in order to put in evidence possible patterns and regularities
- Reduce the number of distinct values in an attribute, in order to increase the efficiency of the computations
- **Reduce the number of distinct valuess in an attribute, in order to put in evidence possible patterns and regularities**
- Reduce the range of values of a numeric attribute, to make all the attributes more comparable

58) Given the following definitions: TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives. Which of the formulas below computes the recall of a binary classifier?

- **TP / (TP + FN)**
- TN / (TN + FP)
- TP / (TP + FP)
- (TP + TN) / (TP + FP + TN + FN)

59) The Information Gain is used to

- **select the attribute which maximises, for a given training set, the ability to predict the class value**
- select the attribute which maximises, for a given test set, the ability to predict the class value
- select the attribute which maximises, for a given training set, the ability to predict all the other attribute values
- select the class with maximum probability

60) Which is the main reason for the MinMax scaling (also known as "rescaling") of attributes?

- Change the distribution of the numeric attributes, in order to obtain gaussian distributions
- Map all the nominal attributes to the same range in order to prevent the values with higher frequency from having prevailing influence
- Remove abnormal values
- **Map all the numeric attributes to the same range, in order to prevent attributes with higher range from having prevalent influence**

61) Which of the following is a base hypothesis for a bayesian classifier?

- **The attributes must be statistically independent inside each class**
- The attributes must have negative correlation
- The attributes must be statistically independent
- The attributes must have zero correlation

62) When developing a classifier, which of the following is a symptom of overfitting?

- The error rate in the test set is much smaller than the error rate in the training set
- The precision is much greater than the recall
- The error rate in the test set is more than 30%
- **The error rate in the test set is much greater than the error rate in the training set**

63) With reference to the total sum of squared errors and separation of a clustering scheme, which of the statements below is true?

- They are two ways to measure the same thing
- They are strictly correlated, if, changing the clustering scheme, one increases, then the other does the same
- It is possible to optimise them (i.e. minimise SSE and maximise SSB) separately
- **They are strictly correlated, if, changing the clustering scheme, one increases, then the other decreases**

64) How can we measure the quality of a trained regression model?

- **With a formula elaborating the difference between the forecast values and the true ones**
- With a confusion matrix
- Counting the number of values correctly forecast
- With precision, recall and accuracy

65) What is the coefficient of determination R2?

- **Provide an index of goodness for a linear regression model**
- Measure the amount of error in a linear regression model
- Measure the amount of error in a regression model
- An index of goodness for a classification model

66)What measure is maximised by the Expectation Masimisation algirithm for clustering?

Scegli un'alternativa:

- **The likelihood the distributions, defined by the parameters found, given the data available**
- The likelihood of an example
- The support of a class
- The likelihood of an attribute, given the class label

67) What is the difference between classification and regression?

- **Classification has a categorical target, while regression has a numeric target**
- Classification is a supervised activity. while regression in unsupervised
- Classification can make errors, while regression is always exact
- Classification can have a numeric or categorical target, while regression has always a categorical target

68) When is polynomial regression appropriate?

- **When the relationship between the predicting variable and the target cannot be approximated as linear**
- When the target values are not linearly separable
- When there is more than one predicting attribute
- When it is necessary to project the data into a dimensional space

69) How is the number of clusters determined in agglomerative clustering?

- By assigning each data point to a cluster based on the final structure obtained from the dendrogram.
- **By cutting the dendrogram at a certain height.**
- By merging the closest clusters according to the defined distance metric.
- By computing the pairwise similarity or distance between all clusters.

70) In data preprocessing, which of the operations below performs aggregation?

- Combination of two attributes in order to obtatn a more general attribute with smaller variability
- **Reduction of the number of distinct values of an attribute. in order to reduce its variability**
- **Reduction of the numbet of rows of the dataset by applying a groupinq on one or more attributes and computing the measures as the results of aggregation functions**
- Combination of two or more datasets in Order to obtatn more Information

71) Match specific situations of data to the most appropriate pre-processing activity

- The numeric attributes have very diverse ranges of values : **Rescaling**
- An attribute has a very skewed distribution (e.g. most of the values concentrated in a small range, and a very long tail of outliers: **Non linear transformations**
- A categorical attribute has a too large number of distinct values: **Aggregation**
- A nominal attribute needs to be used as predicting when using the sklearn library: **OneHot Encoding**

72) What is the first step in agglomerative clustering? *(NON VERIFICATA)*

Scegli un'alternativa:

- Cutting the Dendrogram: Decide on the number of clusters by cutting the dendrogram at a certain height
- Pairwise Cluster Similarity Calculation: Compute the pairwise similarity or distance between all clusters
- **Initialization: Start with each data point as a singleton cluster**
- Cluster Assignment: Assign each data point to a cluster based on the final structure obtained from the dendrogram

73) Which of the following situations could suggest the use of MinMax scaling? (also known as "rescaling')

- **There are numeric attributes with very diverse value ranges**
- There are attributes with very skewed data distributions
- There are attrtbutes with a very large number of missing values
- There is a numeric attribute with values spread in a very large range